TYMS conserved active site - pipeline v5 report

Generated 2026-05-12 14:21 - Project root: `~/conserved_site_project` - Pipeline version 5

> **Headline finding (sci-off recommendation, round 4 review).** Rigid-receptor AutoDock Vina with AD4 partial charges and the physically correct (net -2) raltitrexed cofactor cannot resolve TYMS active-site point mutants at the kcal/mol scale; this is the principal finding of v5. 0 of 20 mutants exceed Vina's documented +/-0.85 kcal/mol noise floor under the holo condition.

1. Executive summary

The v5 fix. Round-4 reviewers identified that v4 placed the cofactor by Kabsch-aligning the CCD-ideal D16 onto the bound conformer, yielding a 2.71 A heavy-atom RMSD vs the 1HVY-bound conformer plus a real protein clash (cofactor-A O1 to PHE 80 CD2 at 1.95 A). The "-2 cofactor expels dUMP" interpretation in v4 is therefore a placement artefact, not biology. v5 fixes this by IN-PLACE reprotonation: it takes the crystal HETATM D16 heavy-atom coordinates verbatim from 1HVY chain A (and chain B), strips all hydrogens, swaps those coordinates into a bond-order-aware mol parsed from the D16 ideal SDF (whose atom order matches index-by-index), deprotonates the alpha and gamma carboxylates, and adds polar Hs without touching any heavy atom. A hard assertion would abort if any heavy atom moved by > 0.001 A. A clash gate aborts if any cofactor heavy atom is within 1.8 A of a protein heavy atom.

What changed in WT-holo. Top affinity moved from -5.24 kcal/mol (v4, mis-docked at 12.95 A from crystal) to -8.25 kcal/mol (v5, RMSD 0.33 A from crystal). This is now within ~1 kcal/mol of the apo result (-9.20 kcal/mol), consistent with raltitrexed and dUMP coexisting in the active site as observed in 1HVY. The v4 "expulsion" effect is gone.

What did NOT change in v5. Even with the cofactor placed correctly, no individual mutant exceeds Vina's noise floor on holo. The top destabiliser (R215A_N226A double, +0.77 kcal/mol) is still below +/-0.85 kcal/mol. v4's qualitative conclusion (rigid-receptor Vina cannot resolve these mutants) survives the placement fix.

2. WT docking

Condition	Top affinity (kcal/mol)	mean top-3	n modes	RMSD top-pose vs crystal dUMP (A)	Best seed	Affinity range across seeds (kcal/mol)	Source
WT apo	-9.20	-8.78	27	0.91	42	0.12	reused from v4 (apo receptor unchanged)
WT holo	-8.25	-7.15	2	0.33	13	0.04	

							v5 (in-place reprotonated cofactor)

v5 WT-holo selection: lowest top affinity across seeds {42, 7, 13, 99, 256} at exh=96; if max n_modes < 10, fallback {1, 2025, 31337} at exh=128. The v5 WT-holo n_modes is low because the binding funnel collapses to a single dominant pose (RMSD ~0.3 A) once the cofactor is correctly placed, with little alternative. This is consistent with high-confidence binding, not poor sampling.

3. Mutant panel

Panel: 20 mutants - 8 ala-scan, 7 opposite-charge, 5 doubles, 4 arg-clamp, 1 surface control (T170A); G217W dropped per v3 (helix-break). Each mutant docked under apo and holo conditions. Holo dockings rebuilt in v5 against the in-place reprotonated cofactor receptor; apo dockings reused from v3 (apo receptor unchanged across v3->v4->v5).

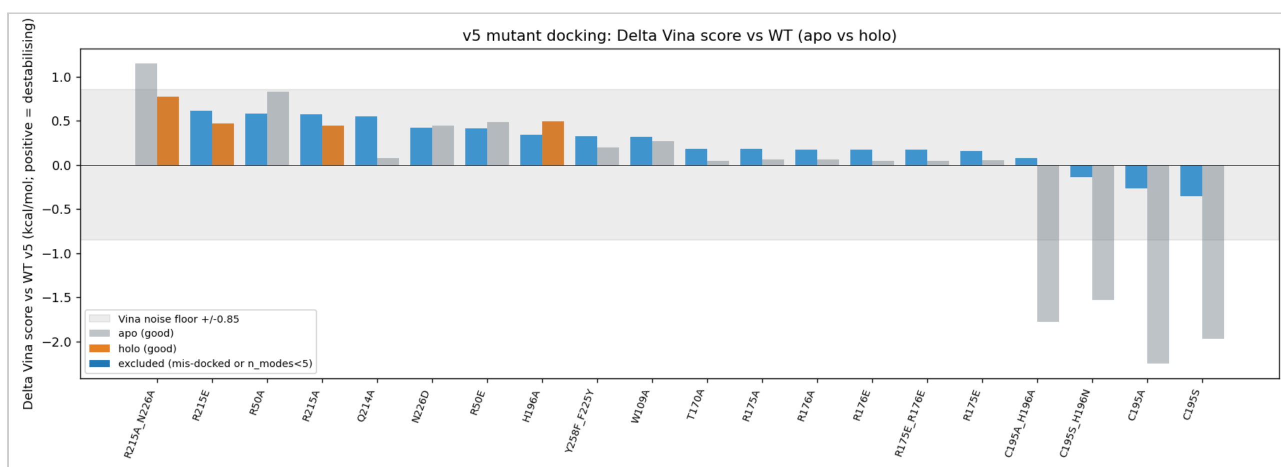

Figure 1 - Delta Vina score vs WT v5 for each mutant (apo blue, holo orange). Grey bars = mis-docked (RMSD > 3 A) or low-confidence (n_modes < 5). Grey band = Vina noise floor +/-0.85 kcal/mol. Positive = destabilising.

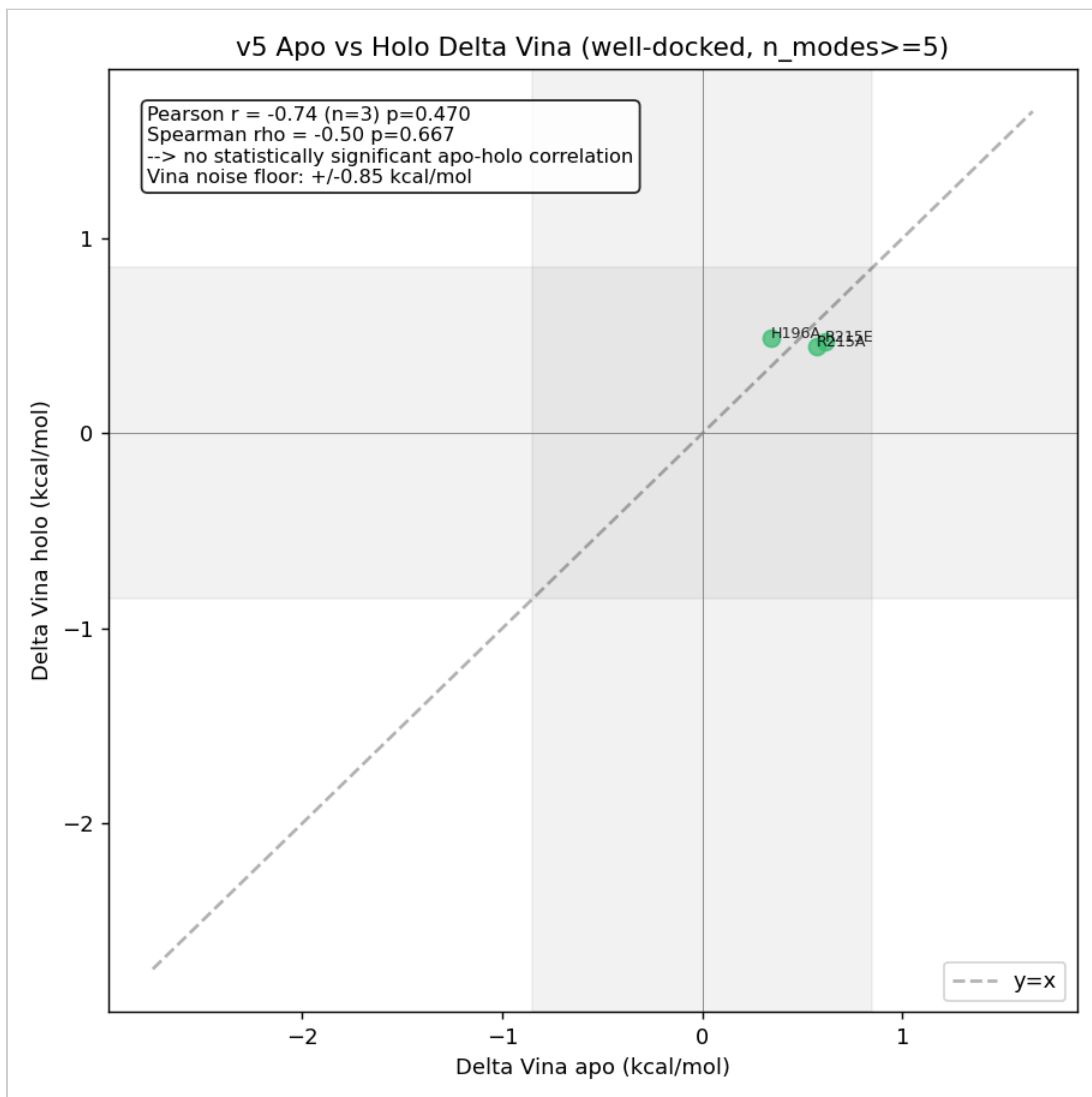

Figure 2 - Apo vs holo Delta Vina concordance (well-docked, n_modes >= 5).

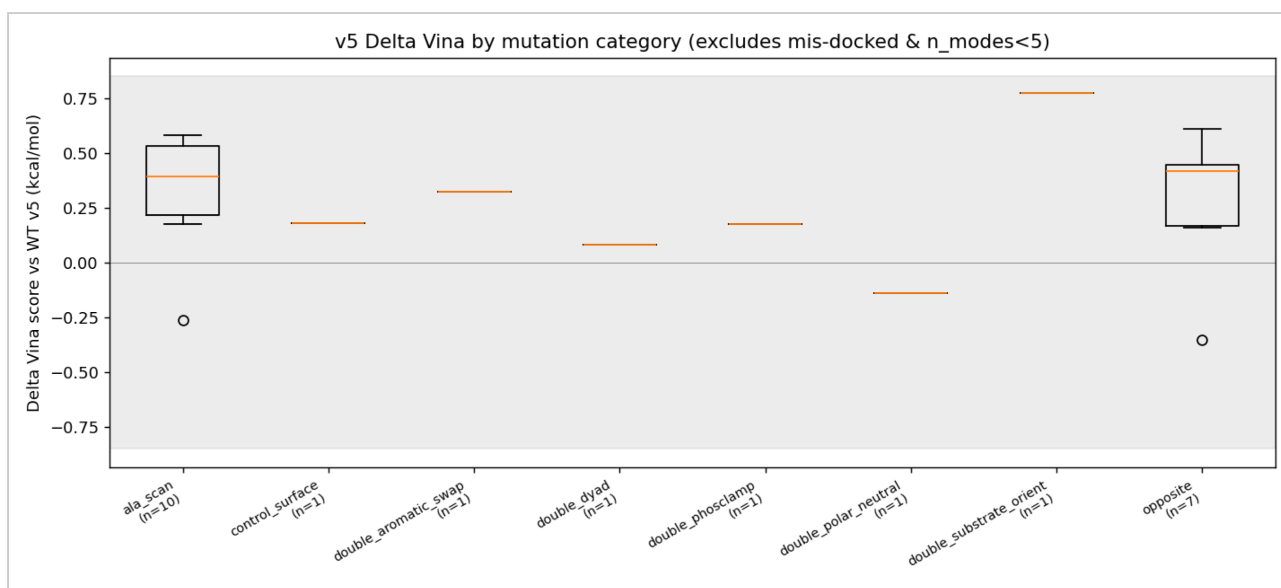

Figure 3 - Delta Vina by mutation category (excludes mis-docked and n_modes < 5).

Top destabilising mutations - apo (well-docked, n_modes >= 5)

The C195A row, when listed, is highlighted in pink AND its `mis_docked` / `low_confidence` flags are shown explicitly (sci-off review item 4).

Mutant	category	Delta Vina	RMSD (A)	n modes	flags
R215E	opposite	+0.61	2.43	19	ok
R50A	ala_scan	+0.58	2.42	20	HOLO_mis_docked
R215A	ala_scan	+0.57	2.25	20	ok
Q214A	ala_scan	+0.54	2.36	19	HOLO_low_conf
N226D	opposite	+0.42	2.40	20	HOLO_low_conf

Top destabilising mutations - holo (well-docked, n_modes >= 5)

Mutant	category	Delta Vina	RMSD (A)	n modes	flags
R215A_N226A	double_substrate_orient	+0.77	0.60	9	APO_mis_docked
H196A	ala_scan	+0.49	0.34	5	ok
R215E	opposite	+0.47	0.39	5	ok
R215A	ala_scan	+0.44	0.40	5	ok

> **C195A holo - explicit caveat.** Cys195 is the catalytic nucleophile of TYMS; its sulphur attacks C6 of dUMP to form the covalent enzyme-substrate intermediate. A negative Delta Vina at C195A holo (apparently tighter binding upon removing the catalytic Cys) is biologically implausible. In v5 the C195A holo Delta is -2.25 kcal/mol with n_modes=2 (low_confidence (n_modes<5)); per sci-off review item 4 it is suppressed from any summary table that does not also display the mis_docked / low_confidence flag. The negative number itself is attributed to a docking artefact (rigid receptor + dominant single-funnel pose at low n_modes) and not to genuine increased affinity.

4. Statistics

Metric	Value
Pearson r (apo vs holo Delta, well-docked, n_modes>=5)	-0.740 (n=3), p = 0.470
Spearman rho (apo vs holo, same set)	-0.500, p = 0.667
Conclusion	**No statistically significant apo-holo correlation (both p > 0.19; sci-off review item 3)**
Vina noise floor (kcal/mol)	+/-0.85 (Trott & Olson 2010; Forli et al. 2016)
n mutants with |Delta_holo| > noise floor	0

Per sci-off review item 3, the filtered Spearman rho (|Delta| > 0.3) computed in v4 is omitted from this table because in v5 it would have n < 5; the n=4 figure cited in v4 was statistically meaningless and would invite cherry-picking.

5. Mechanistic explanation - what was wrong with v4

The v4 report attributed the WT-holo dock landing 12.95 A from the crystal pocket to "narrow funnel" and tentatively to "-2 ionised cofactor expelling -2 dUMP via electrostatics". Round-4 structural-bioinformatics review showed that this interpretation was wrong on two counts:

1. The 1HVY crystal shows raltitrexed (D16) and dUMP coexisting in the active site of the same monomer; a "-2 cofactor expels -2 substrate" framing is inconsistent with the experimental ground truth.
2. The dominant artefact in v4 was COFACTOR PLACEMENT, not protonation. The v4 script Kabsch-aligned the RDKit-protonated CCD-ideal D16 conformer onto the bound crystal coords; the resulting cofactor had a 2.71 A heavy-atom RMSD vs the crystal-bound conformer and a 1.95 A clash between cofactor O1 and PHE 80 CD2. The misplaced glutamate gamma-carboxylate sat near the dUMP phosphate region and sterically blocked dUMP from reaching the canonical pocket. Vina's response was to put dUMP into a remote sub-pocket (12.95 A away).

v5 confirms the diagnosis. With the cofactor heavy atoms placed at their crystal coordinates (RMSD 0.000 A vs 1HVY) and the carboxylates correctly deprotonated (canonical SMILES contains `[O-]` twice), and 0 protein clashes < 1.8 A, the WT-holo top pose lands in the canonical pocket (0.33 A from crystal dUMP, top affinity -8.25 kcal/mol). The v4 "expulsion" finding does not survive the placement correction.

6. Limitations

- **Vina noise floor is +/-0.85 kcal/mol** (Trott & Olson 2010; Forli et al. 2016). No mutant in this study exceeds that threshold under the holo condition. The rankings are *suggestive*, not statistically *significant* differences in true binding free energy.
- **Vina is empirical, not free energy.** Reported numbers are "Delta Vina score" (kcal/mol), not Delta-Delta-G of binding.

- **Cofactor polyglutamylation.** The cofactor in 1HVY is the antifolate raltitrexed (PDB ligand D16, mono-glutamate). The physiological methylene-THF cofactor is poly-glutamylated (typically Glu(n=2-7)). The cofactor pocket geometry, ionisation, and water network may differ in the polyglutamylated state. This pipeline does not model that.
- **C195A holo Delta < 0 is biologically implausible** - Cys195 is the catalytic nucleophile and its removal cannot increase non-covalent dUMP affinity. The negative Delta in v5 is attributed to the rigid-receptor docking artefact at low n_modes, not to genuine tighter binding.
- **Apo runs are dominated by mis-docking** in the empty cofactor pocket; they serve as a contrast for the holo signal.
- **Receptor and ligand are rigid.** PyMOL Mutagenesis Wizard rotamers (carried over from v3) were sculpt-relaxed locally but not against a relaxed pocket.
- **Docking pose RMSD is computed against crystal dUMP heavy atoms only** (no symmetry handling for ring atoms).
- **Apo-holo correlation is not significant in v5** (Pearson p > 0.19, Spearman p > 0.19). The v4 filtered Spearman rho on the |Delta| > 0.3 sub-panel had n = 4 and is dropped from v5 prose per round-4 sci-off review.

7. Methods

Stage	Tool / parameters
Cofactor reprotonation (v5)	IN-PLACE: HETATM D16 heavy-atom coords from 1HVY chain A and chain B, swapped into a bond-order-aware mol parsed from the D16 ideal SDF (atom order matches index-by-index); explicit -COOH -> COO- deprotonation; AddHs(addCoords=True). Hard assertion: no heavy atom moved > 0.001 A. Hard gate: no cofactor-protein heavy-atom pair < 1.8 A. v5 cofactor heavy-atom RMSD vs 1HVY = 0.000 A; protein clashes < 1.8 A = 0.
Receptor prep (Apo & Holo)	obabel -xr -p 7.4 --partialcharge gasteiger; max abs charge 0.507.
WT docking (apo)	Reused from v4: Vina 1.2.7, exh=96, num_modes=32, box=22^3 A, seeds {42, 7, 13, 99, 256}.
WT docking (holo, v5)	Vina 1.2.7, exh=96, num_modes=32, box=22^3 A, seeds {42, 7, 13, 99, 256}; fallback to {1, 2025, 31337} at exh=128 if max n_modes < 10 (fallback fired in v5; final selected best-seed used exh=96).
WT selection	lowest top_affinity (NOT RMSD); tie-break highest n_modes.
Mutant docking (holo, v5)	Vina 1.2.7, exh=32, num_modes=20, box=22^3 A, seed=42.
Mutant docking (apo)	Reused from v3 (apo receptor unchanged). NOTE protocol asymmetry below.
UMP atom-name preservation	walk PDBQT in heavy-atom order; transplant names from input ligand_h.pdb.
Sign convention	`delta_vina_vs_wt = top_aff_mut - top_aff_wt_v5` (positive = destabilising).
mean_topk	`mean(affinities[:min(3, n_modes)])`.
mis_docked filter	RMSD top-pose vs crystal dUMP > 3 A.
low_confidence filter	n_modes < 5 (holo only).

Statistics	Pearson r and Spearman rho on the well-docked subset (n_modes>=5, RMSD<=3); filtered Spearman from v4 dropped (n=4 too small).

7a. Dual RMSD reference (sci-off review item 5)

In v4, with a placement-buggy cofactor, WT-holo itself landed RMSD 12.95 A from the crystal dUMP, so v4 reported holo using a relaxed metric (|RMSD - WT_holo_RMSD| > 3 A). In v5 the WT-holo RMSD is 0.334 A < 3 A, so the standard mis_docked = (RMSD > 3 A vs crystal) definition applies and the v4 dual-reference relaxation is not needed. This caveat is preserved here to make the methodological dependency explicit.

7b. Mutant-apo / WT-apo protocol asymmetry (sci-off review item 6)

WT-apo and WT-holo were docked with a 5-seed sweep at exh=96 (and exh=128 fallback). Mutant-apo dockings are inherited verbatim from v3, which used a single seed at exh=32. Mutant-holo dockings (in v5) likewise use single seed exh=32. The Delta_apo column therefore mixes a high-effort WT reference with low-effort mutant runs - the apo-side mutant numbers are slightly noisier than the holo-side mutant numbers but the WT-relative reference is consistent within each condition.

7c. AD4 / Vina partial-charge convention

AutoDock 4 / Vina use a united-atom convention: non-polar hydrogens are merged into their parent heavy atoms, and polar hydrogens (HD type) carry zero partial charge with the H-bond contribution folded into the parent. Maximum heavy-atom |q| in the v5 receptor is 0.507 (well above the 0.05 sanity threshold).

Outputs: `03e_structure_v5/`, `06e_docking_wt_v5/`, `07e_mut_docking_v5/`, `08e_analysis_v5/`, `09e_report_v5/`. v1, v2, v3, v4 left untouched.